Applied A.I. Solutions

Data Visualization Techniques

Professor

Daniel Vitaver, EMBA

daniel.vitaver@georgebrown.ca

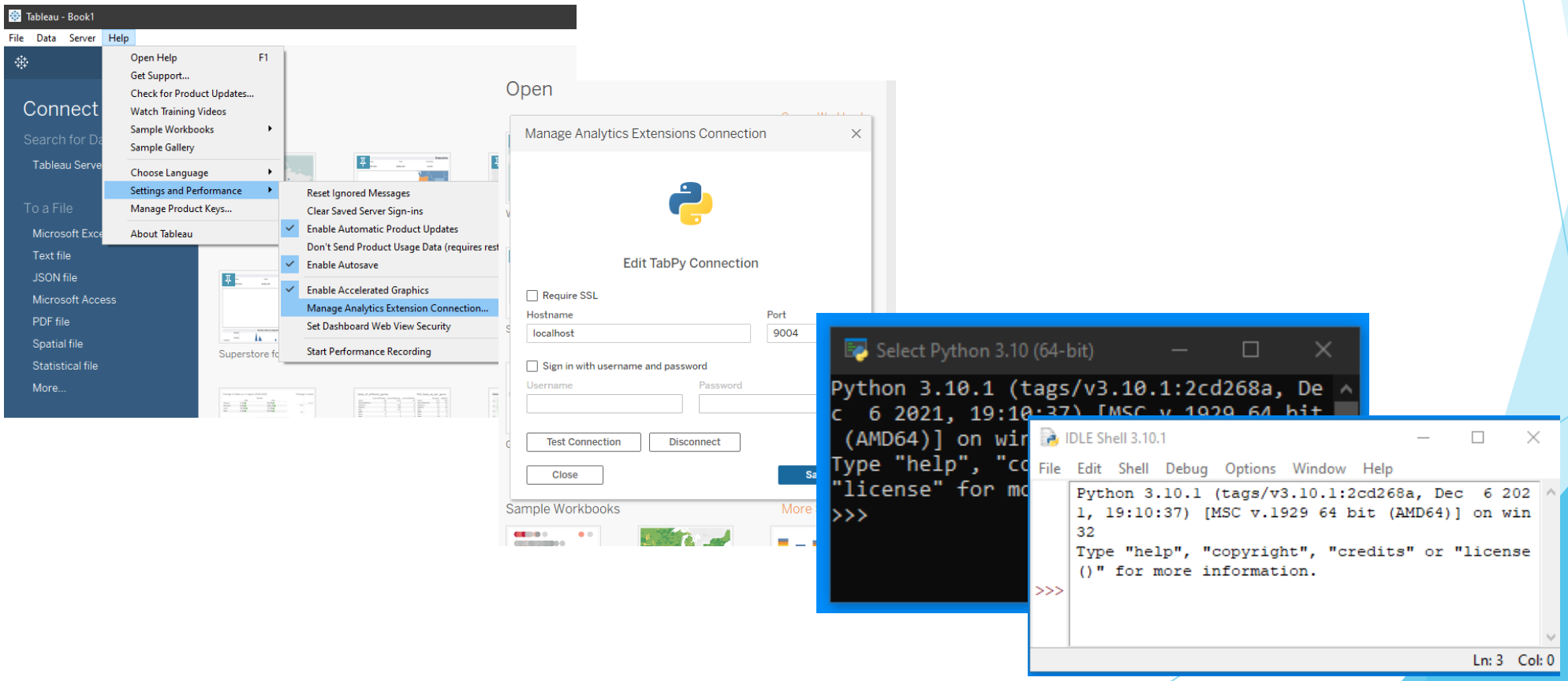
DATA VISUALIZATION TECHNIQUES

Tableau - Python Integration

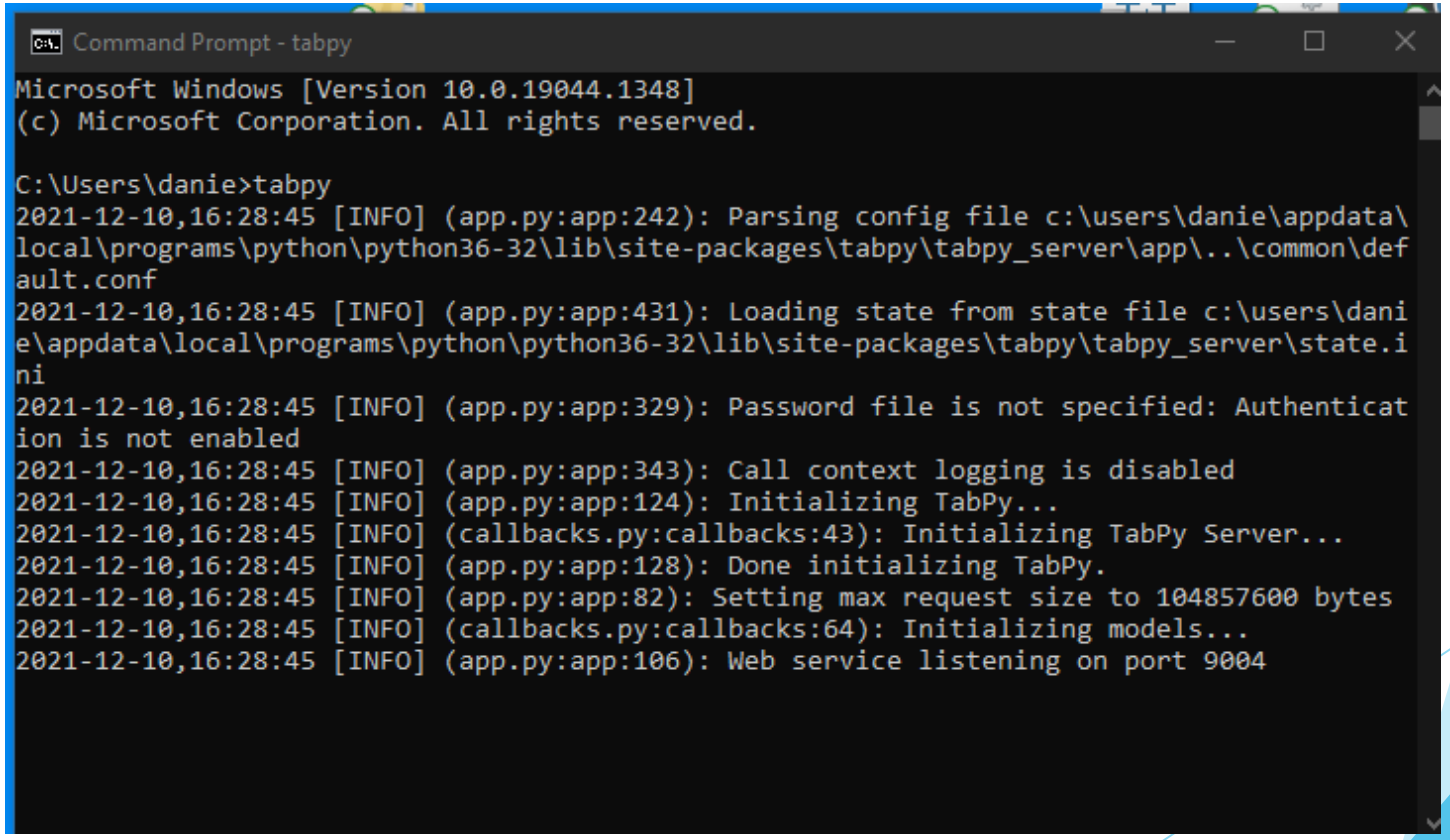
Tableau Python Server - TabPy

- TabPy is an API that enables evaluation of Python code from within a Tableau workbook.
- When you use TabPy with Tableau, you can define calculated fields in Python, leveraging the power of machine-learning libraries right from your visualizations.
- This Python integration in Tableau enables powerful scenarios. For example, **it takes only a few lines of Python code to get data you can explore in using Tableau**.

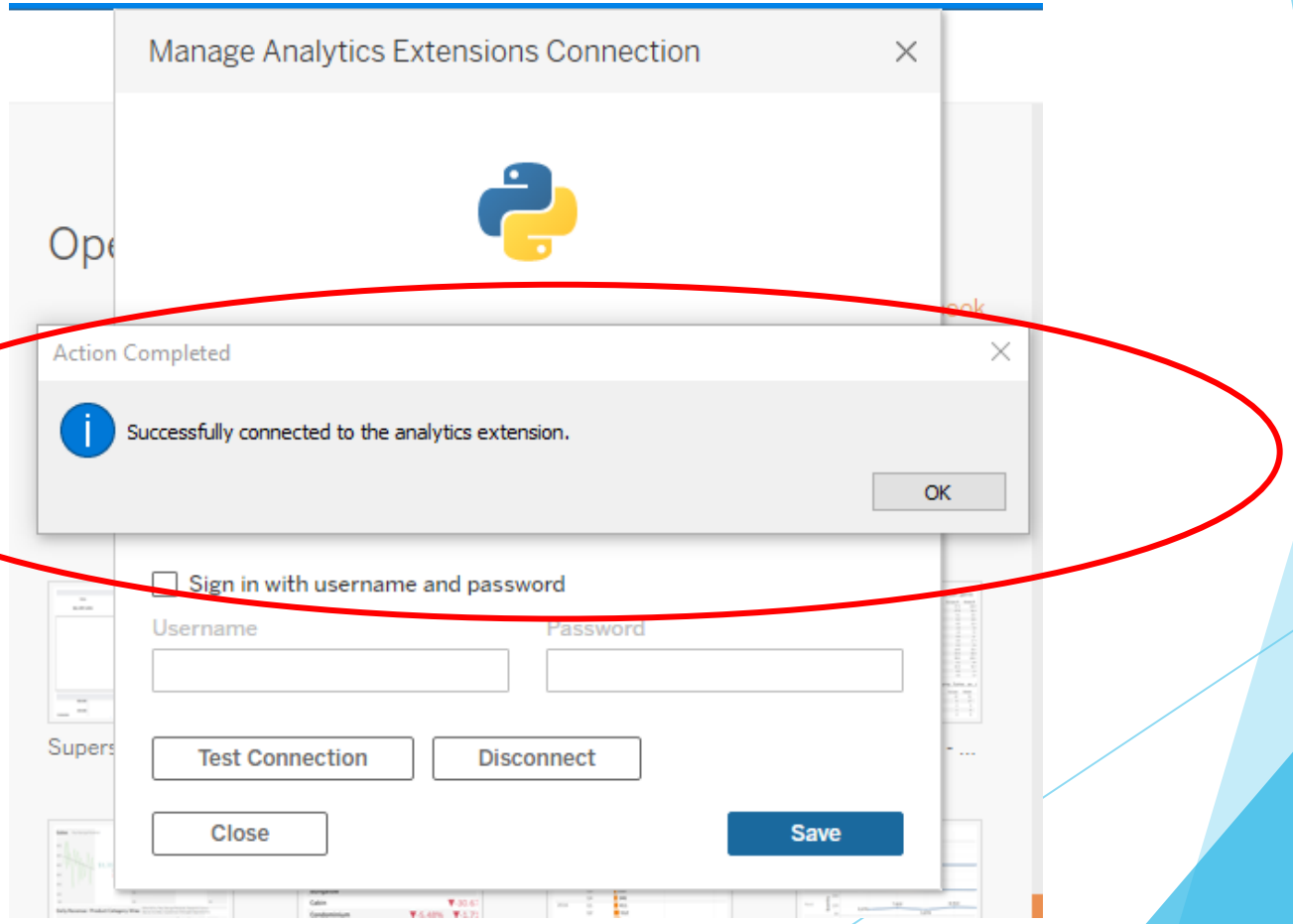
1. **Open Tableau Pro Desktop**
2. **Test Python/TabPy Connection to Tableau** (localhost, 9004)
3. **Run Python**
4. **Run IDLE** (Integrated Development and Learning Environment)



5. Run c:\ Command Prompt
6. From c:\ run *pip install TabPy (install TabPy)*
7. From c:\ run *TabPy*



5. **Leave all the opened environments running**
6. **Go to Tableau Desktop**
7. **Test Python/TabPy Connection to Tableau** (localhost, 9004)
8. **Press OK**
9. **Save**



To verify the installation of TabPy in a web browser type:

http://localhost:9004/

The browser renders

Now you are ready to open Tableau Desktop

GitHub - tableau/TabPy

TabPy Installation Instructions | TabPy (tableau.github.io)

Importing Libraries in Tableau calculated fields

Importing Libraries in Tableau calculated fields

- import pandas as pd
- import numpy as np
- import matplotlib.pyplot as plt
- import seaborn as sns
- %matplotlib inline
- import warnings
- warnings.filterwarnings('ignore')
- from sklearn.preprocessing import LabelEncoder

Python Functions

- To let Tableau, know that the calculations are using TabPy, it must be passed through one of the four python functions
- Tableau aggregated measures like SUM(), ATTR(), MIN(), MAX() **can also be included** in TabPy calculations
- TabPy functions are **always** computed as Table calculations in Tableau

Python Functions

Python functions that are used in the calculated field to indicate TabPy are:

- **SCRIPT_REAL**: returns an output of type real from the given calculation
- **SCRIPT_INT**: returns an output of type integer from the given calculation
- **SCRIPT_STR**: returns an output of type string from the given calculation
- **SCRIPT_BOOL**: returns an output of type Boolean from the given calculation

Whenever any of these functions are used in a calculated field the calculation is passed directly to a **running TabPy external service instance**.

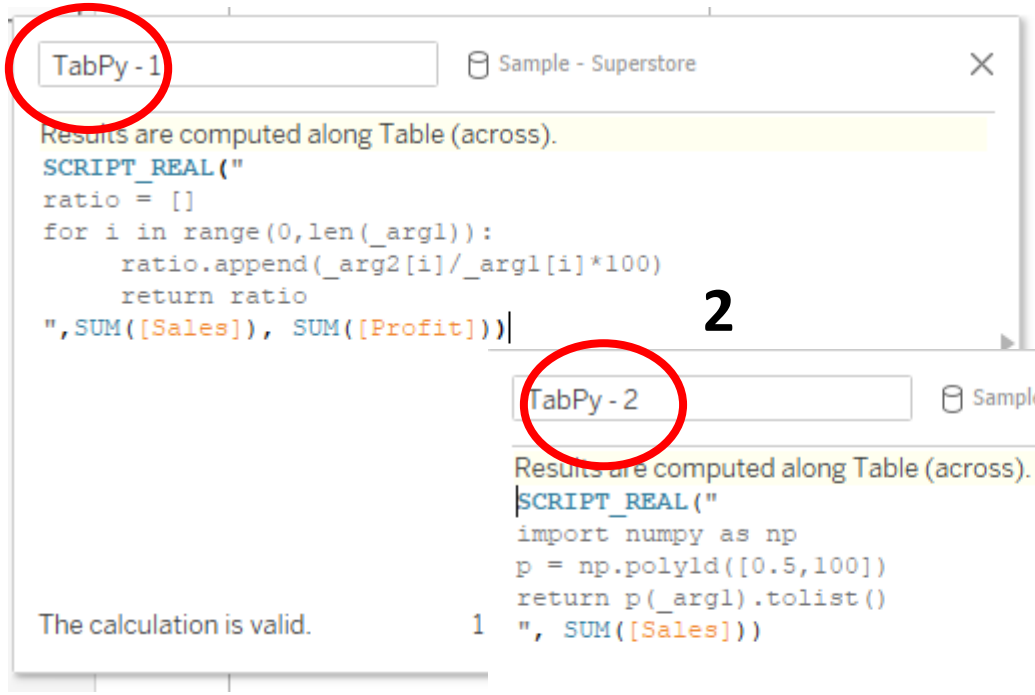
TabPy Example for each Python function

SCRIPT_BOOL: Profitable Sales

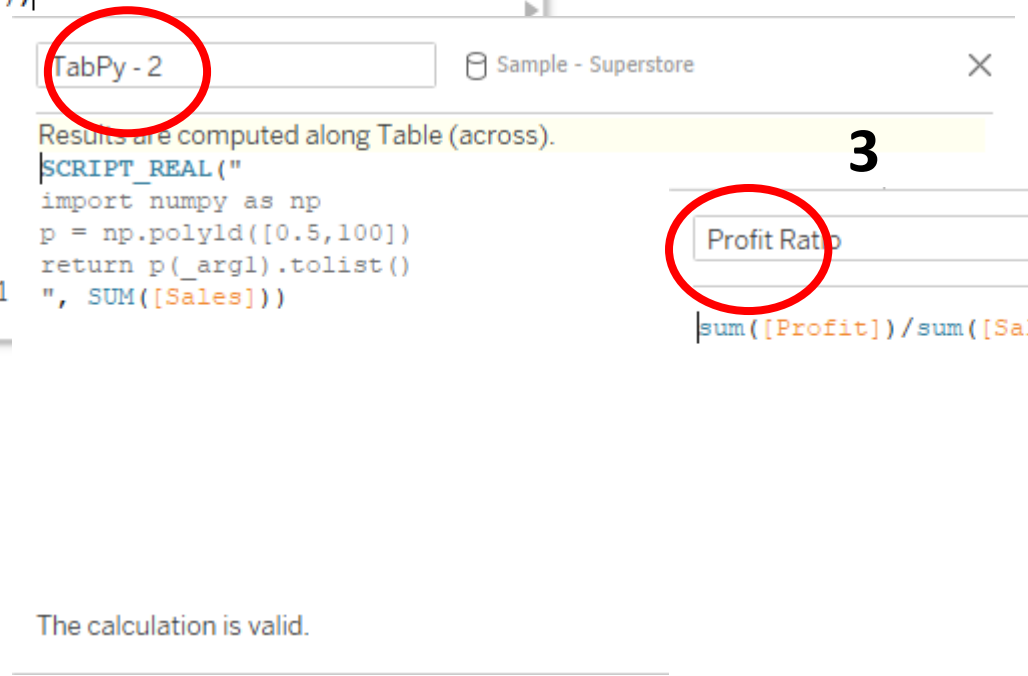
1. We use Sample **Superstore** data to find the Profitable Sales of products
2. Since the profitable Sales can be either true or false, we use **SCRIPT_BOOL**
3. Tableau aggregates data **before** sending it to TabPy using the level of detail of the view.

Python – Numpy library Integration

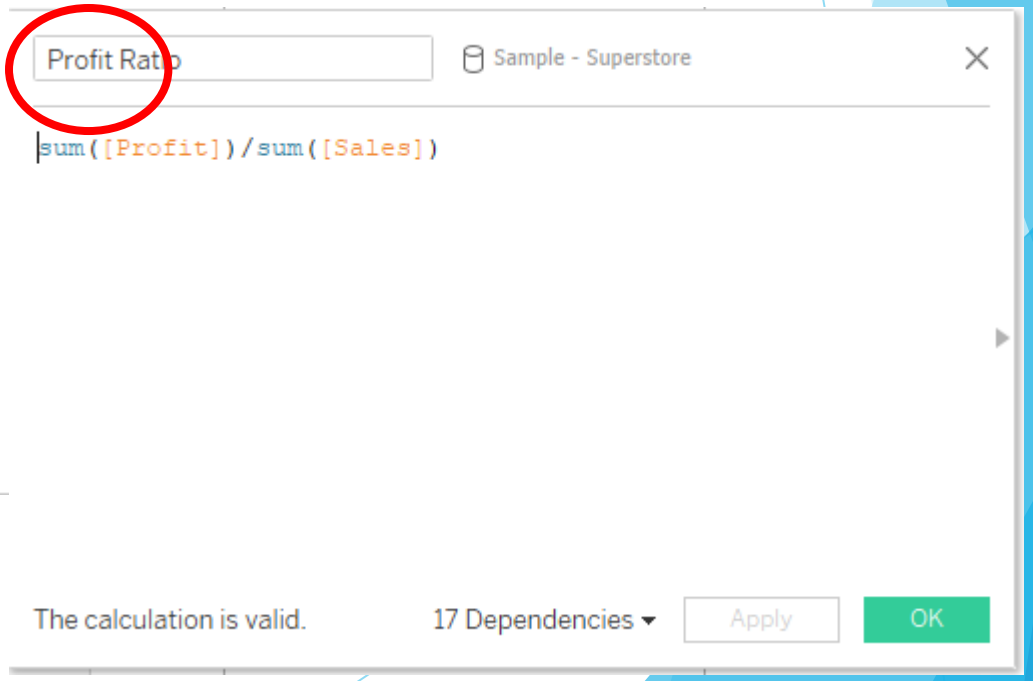
1



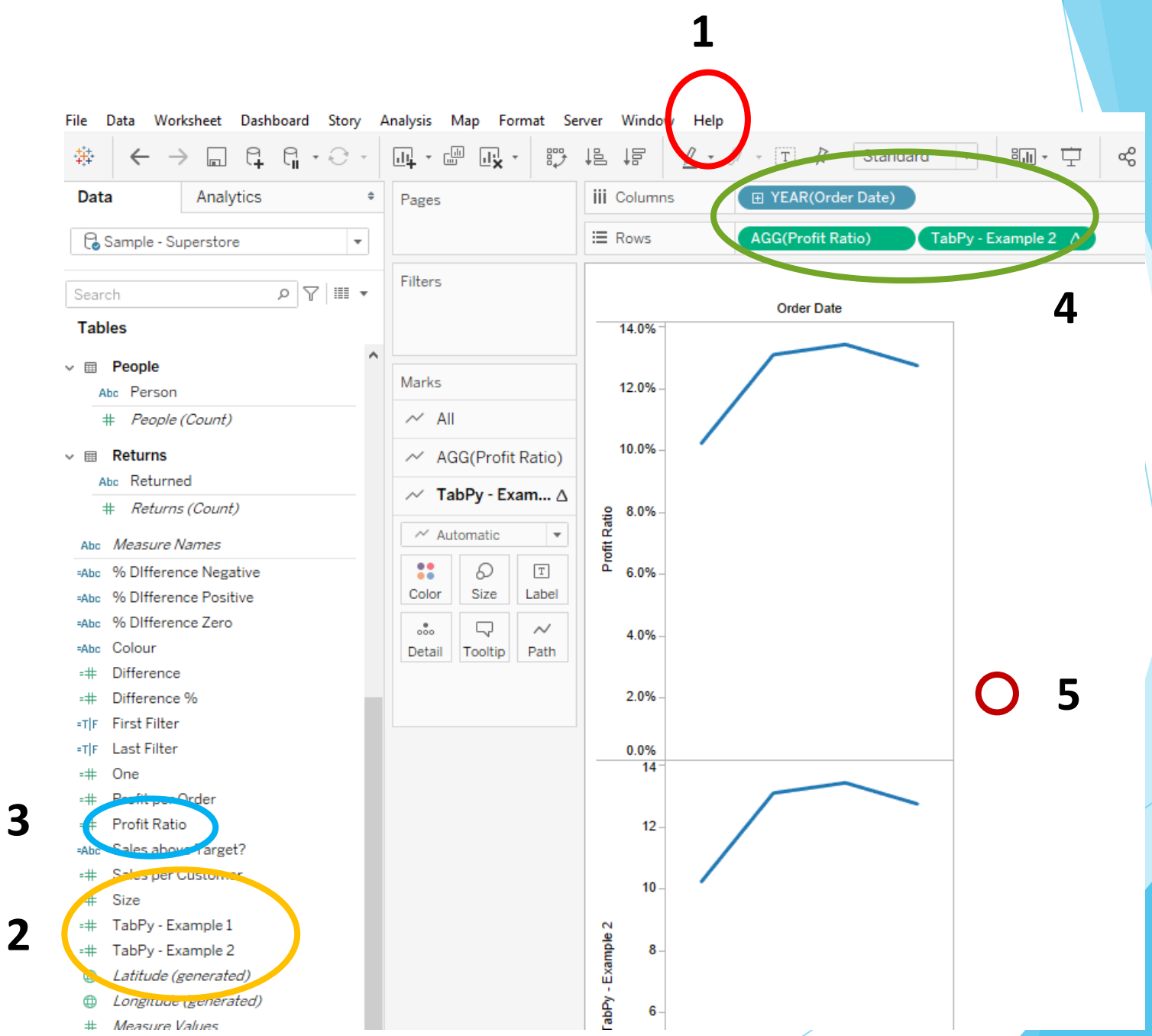
2



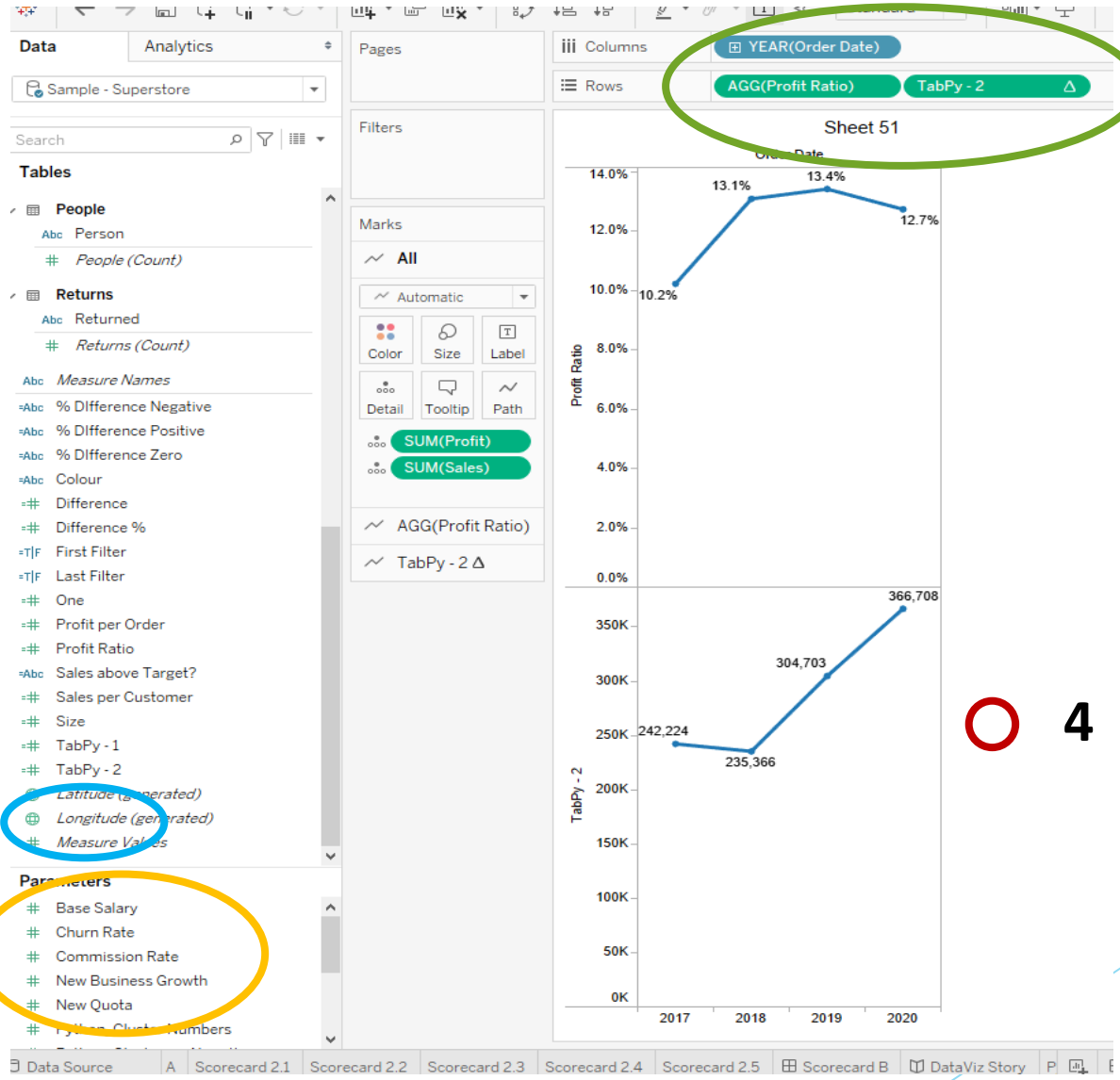
3



Python – Numpy library Integration



Python – Numpy library Integration



1

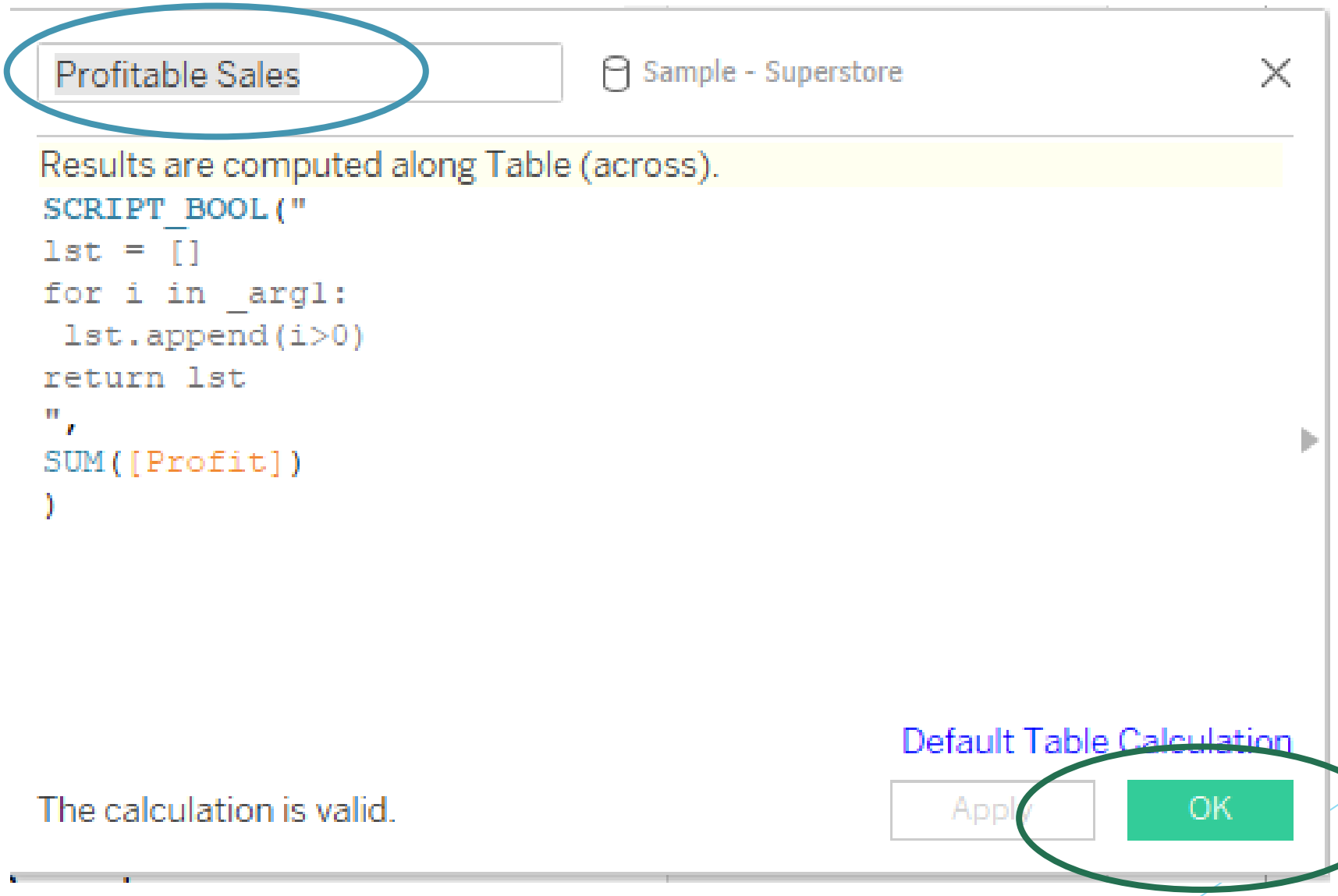
2

3

4

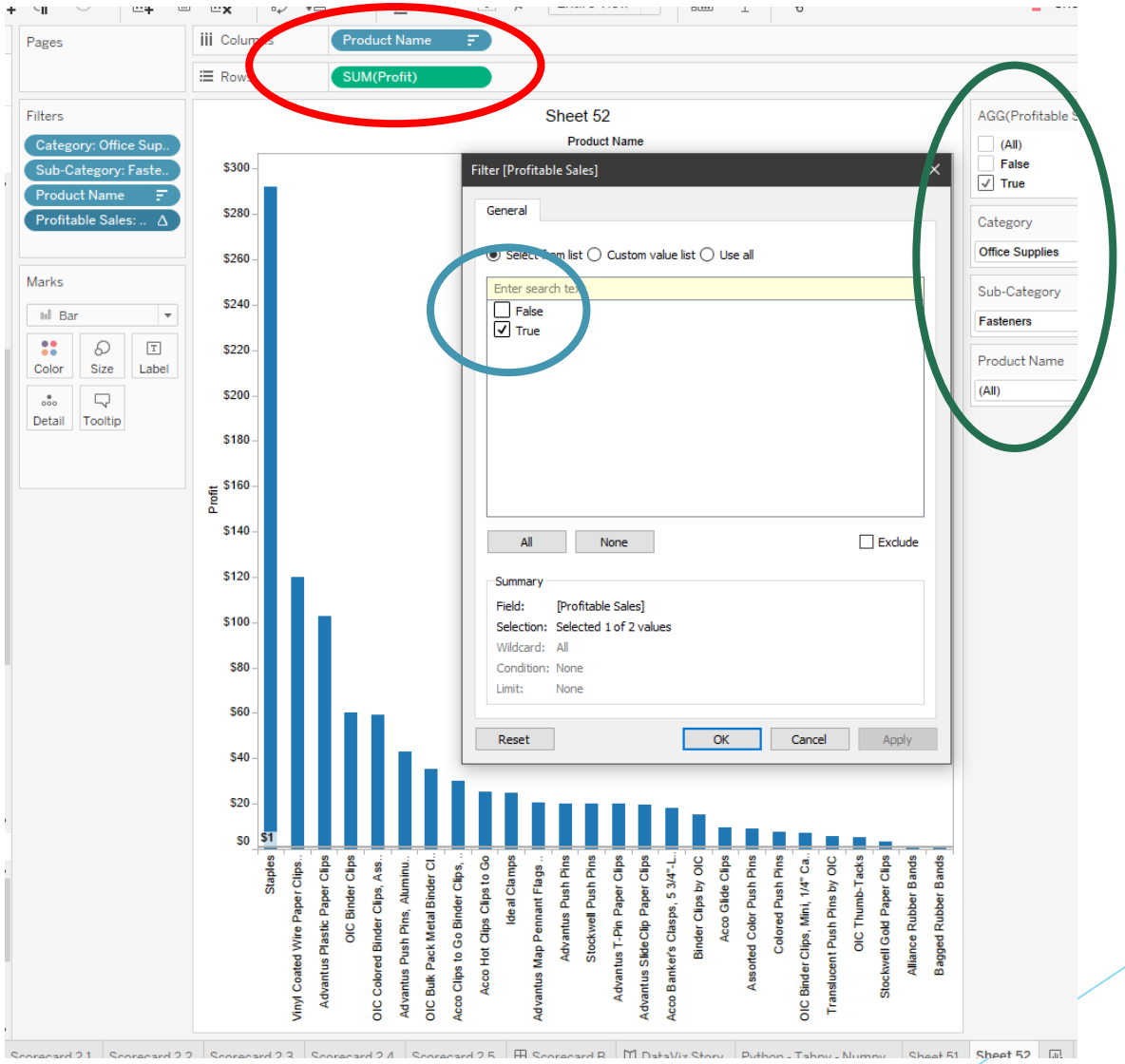
TabPy Example 1

Python calculation of Profitable Sales



See importing Python libraries to Tableau

Python calculation of Profitable Sales



TabPy Example 2

Pearson Correlation Coefficient

- The Pearson coefficient is a type of correlation coefficient that represents the **relationship between two variables that are measured on the same interval or ratio scale.**
- The Pearson coefficient is a measure of the **strength** of the association between two continuous variables
- Pearson correlation coefficient or the bivariate correlation is the measure used to measure the **linear relationship between two variables**.
- Correlation can either be **positive or negative** indicating if a particular variable is impacting another variable positively or negatively.
- **Pearson's correlation should be used only when there is a linear relationship between variables.** It can be a positive or negative relationship, as long as it is significant.

Python calculation of Profitable Sales

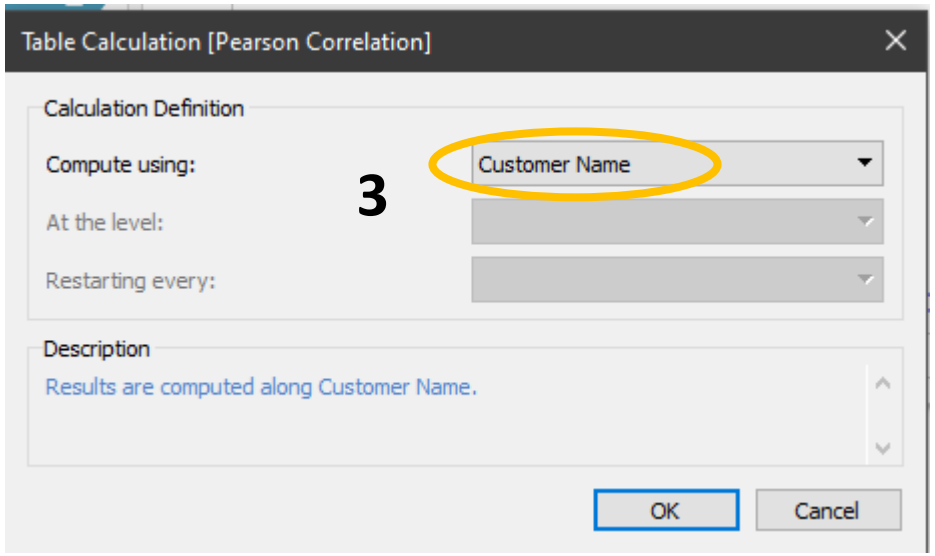
2

Pearson Correlation

Sample - Superstore

Results are computed along Customer Name.

```
SCRIPT_REAL("import numpy as np
return np.corrcoef(_arg1,_arg2)[0,1]",
SUM([Sales]),SUM([Profit]))
```



3

Default Table Calculation

The calculation is valid.

1 Dependency

Apply

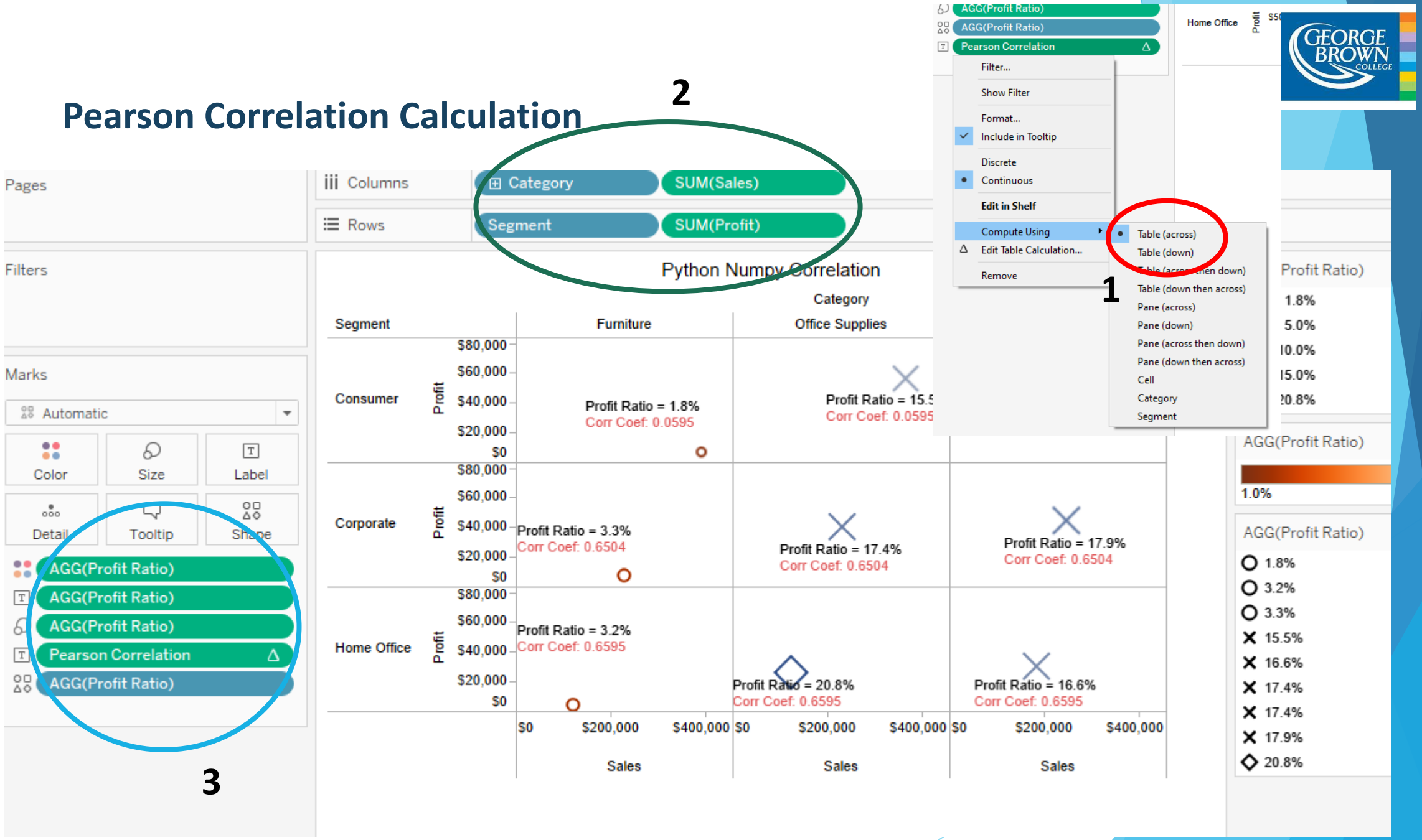
OK

4

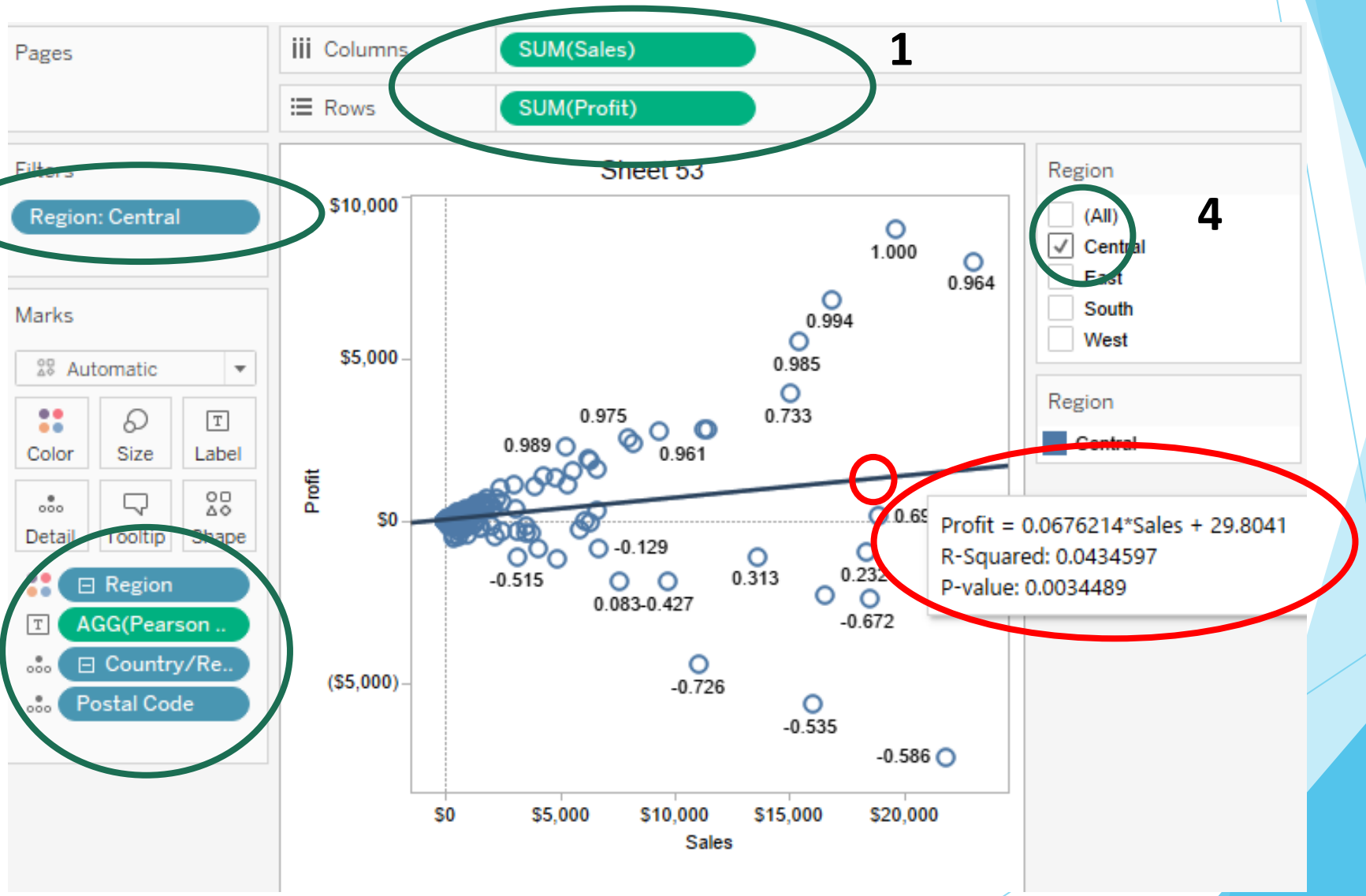
1

See importing Python libraries to Tableau

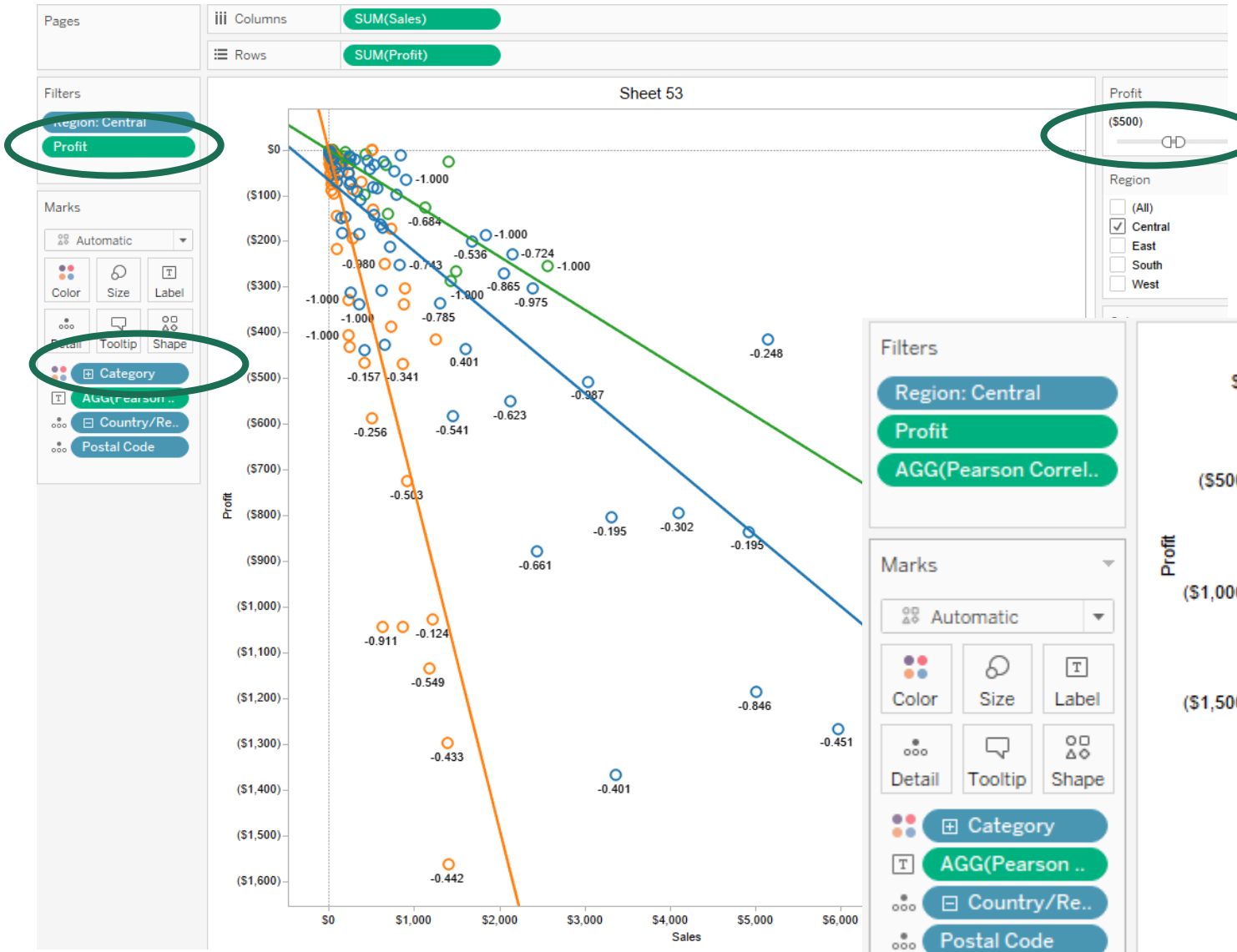
Pearson Correlation Calculation



Correlation Using Tableau Only – Function CORR

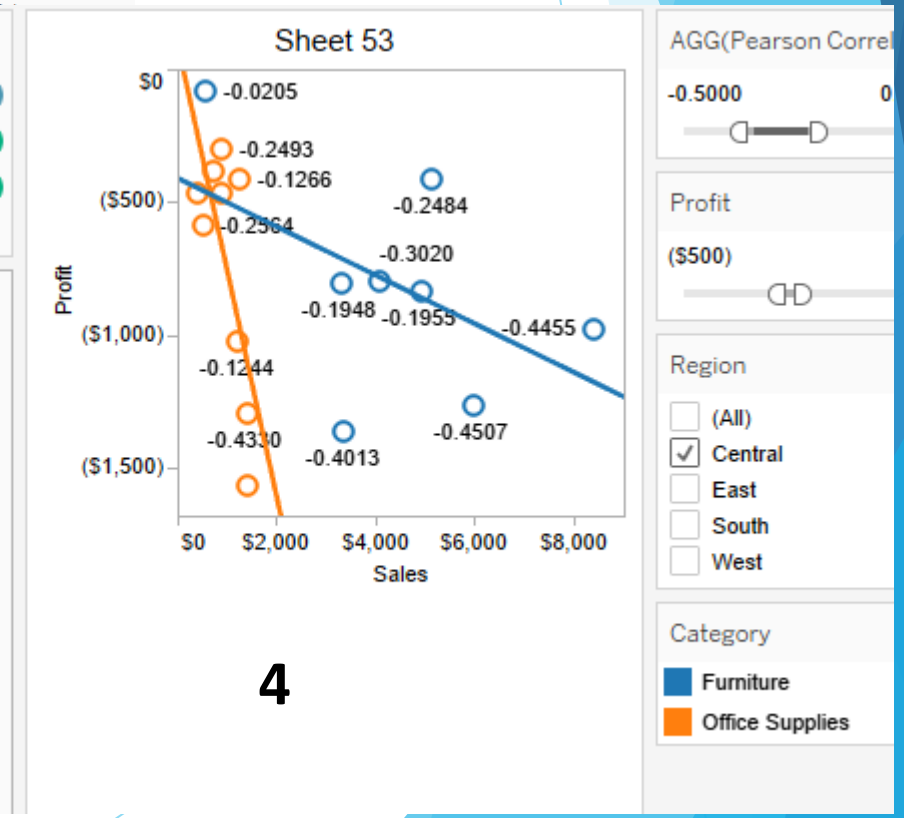


1


Pages
Columns
SUM(Sales)
Rows
SUM(Profit)
Filters
Region: Central
Profit
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Shape
Category
AGG(Pearson ..
Country/Re..
Postal Code
Sheet 53
Profit
($500)
Region
(All)
Central
East
South
West
Profit
Sales

2

3


Filters
Region: Central
Profit
AGG(Pearson Correl..
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Shape
Category
AGG(Pearson ..
Country/Re..
Postal Code
Sheet 53
Profit
Sales
AGG(Pearson Correl
-0.5000
Profit
($500)
Region
(All)
Central
East
South
West
Category
Furniture
Office Supplies

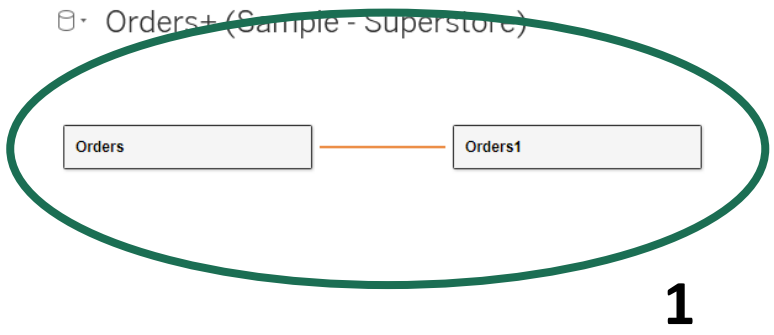
4

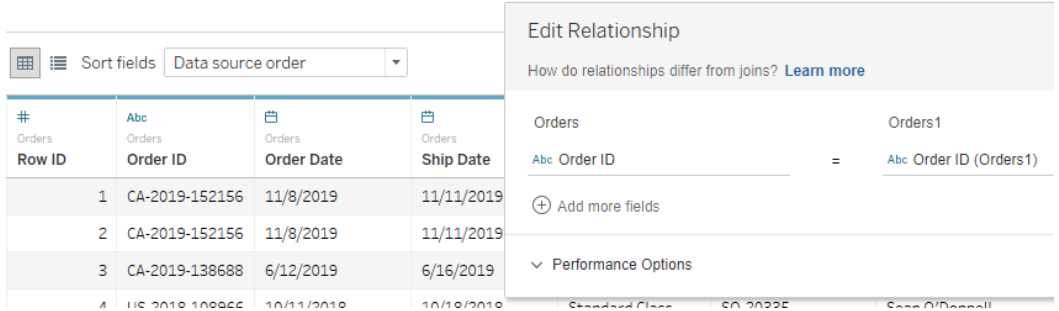
Correlation Matrix Using Tableau Only – Function CORR

https://www.youtube.com/watch?v=vhrmmvAvOfQ

Tableau Custom Charts series: Download Supplemental Materials - Page - SuperDataScience | Machine Learning | AI | Data Science Career | Analytics | Success

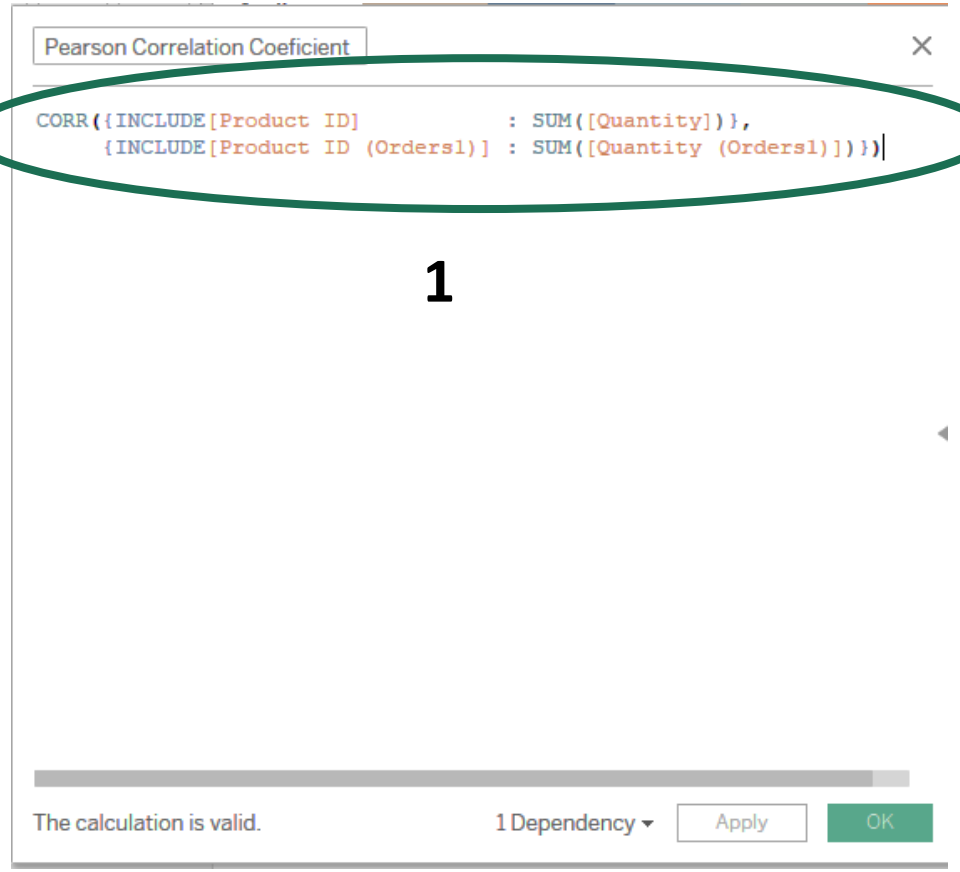
Correlation Matrix Using Tableau Only – Function CORR





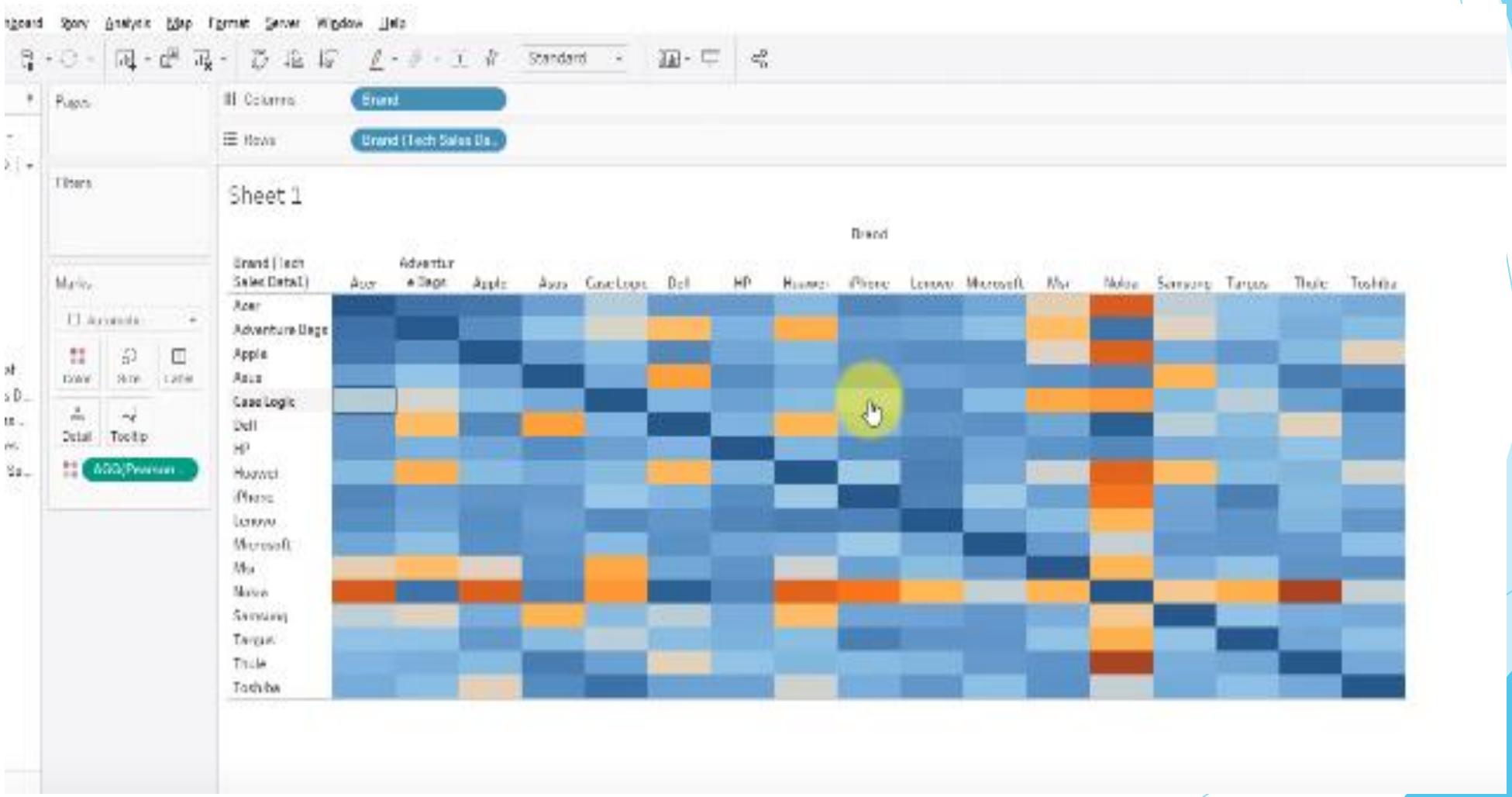
https://www.youtube.com/watch?v=vhrmmvAvOfQ

Correlation Matrix Using Tableau Only – Function CORR



1

Correlation Matrix Using Tableau Only – Function CORR



Correlation: must watch videos

- **Tableau Correlation Matrix**

  https://www.youtube.com/watch?v=vhrmmvAvOfQ

- **How to find the Pearson correlation in Tableau**

  https://www.youtube.com/watch?v=Ns8HulbYnp8

- **NumPy, SciPy, and Pandas: Correlation With Python**

  https://realpython.com/numpy-scipy-pandas-correlation-python/#pearson-correlation-coefficient